PART2　データの分析

目標
- データ分析の方法について理解する。
- 表計算ソフトウェアやプログラムを用いてデータ分析できるようになる。
- データの持つ特性に合わせて分析方法を選択することができるようになる。

Chapter1　データの分析

データの分析手法

　データに基づいた適切な意思決定をするためには，得られたデータが持つ情報をさまざまな側面から明確にし，データの特徴を把握することが重要である。

• 平均値（AVERAGE関数）

　平均値は，データの値の合計をデータの総数で割った値のことである。

　表計算ソフトウェアでは，〔①　　　　　　　　〕関数を使って平均値を求めることができる。

覚えておこう

平均値のようにデータの特徴を要約した数値のことを統計量という。

	A	B	C	D	E	F	G
1	各社の年収(単位万円)						
2		Aさん	Bさん	Cさん	Dさん	Eさん	平均値
3	U社	100	100	200	300	4300	=AVERAGE(B3:F3)
4	J社	400	600	600	800	1100	700

（G3のポップアップ表示：1000 ×）

• 中央値（MEDIAN関数）

　中央値は，データを小さい順に並べたとき，ちょうど中央にある値のことである。

　表計算ソフトウェアでは，〔②　　　　　　　〕関数を使って中央値を求めることができる。

覚えておこう

データ数が偶数の場合は，中央の 2 つの値の平均値を中央値とする。

	A	B	C	D	E	F	G
1	各社の年収(単位万円)						
2		Aさん	Bさん	Cさん	Dさん	Eさん	中央値
3	U社	100	100	200	300	4300	=MEDIAN(B3:F3)
4	J社	400	600	600	800	1100	600

（G3のポップアップ表示：200 ×）

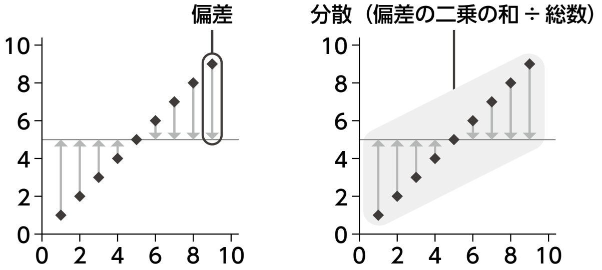
・最頻値（MODE関数）

最頻値は，最も出現する頻度の多い値のことである。表計算ソフトウェアでは，〔③　　　　　　〕関数を使って最頻値を求めることができる。

	A	B	C	D	E	F	G
1	各社の年収(単位万円)						
2		Aさん	Bさん	Cさん	Dさん	Eさん	100 × 最頻値
3	U社	100	100	200	300	4300	=MODE(B3:F3)
4	J社	400	600	600	800	1100	600

・分散と標準偏差（VARP関数とSTDEV.P関数）

分散と標準偏差は，どちらもデータが平均値からどれだけ広い範囲に分布しているかの散らばり具合を表す値のことである。



分散は，偏差の二乗の平均値で，標準偏差は，分散の正の平方根である。

分散
$$\sigma^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n}$$

標準偏差
$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n}}$$

（$x_1 \sim x_n$はデータの値, $\bar{x}$は平均値, nはデータ数）

Google スプレッドシートを用いた表計算では，分散は〔④　　　　　　〕関数, 標準偏差は〔⑤　　　　　　〕関数を使って求めることができる。

	A	B	C	D	E	F	G	H
1	各社の年収(単位万円)							
2		Aさん	Bさん	Cさん	Dさん	Eさん	2728000 × 分散	1651.665826 × 準偏差
3	U社	100	100	200	300	4300	=VARP(B3:F3)	=STDEV.P(B3:F3)
4	J社	400	600	600	800	1100	56000	236.6431913

覚えておこう

偏差は，データの値から平均値を引いた値のことである。

知っておこう

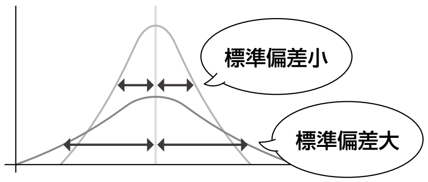
Google スプレッドシートでは VARP 関数だが，Excel では，VAR.P 関数となる。

覚えておこう

成績評価などでよく使われる偏差値は，以下の計算式で求めることができる。

$$偏差値 = 50 + \frac{得点 - 平均値}{標準偏差} \times 10$$

一般的には，データの散らばり具合が少ない程，標準偏差は小さくなる。



重要
分散も標準偏差も，各値が平均値からどれだけ離れているかを表す指標である。それぞれの指標が何を表しているかを理解して使うことが大切である。

- **外れ値による影響**

分析対象のデータの中に，〔⑥　　　　　〕が含まれている場合，分析結果に大きな影響を与える可能性がある。

例えば，先ほどのデータのうち，U 社の E さんの値を 4300 から 1100 に変更すると，1 つの値の影響で平均値・分散・標準偏差が大きく変化する。

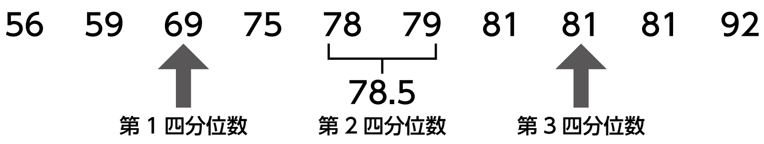
復習
外れ値とは，データの中で，他のものと比べて大きく外れた値のことである。

重要
外れ値は，入力ミスや測定ミスが原因でない場合は，発生原因をよく考察することで，問題の発見や解決の手がかりになる可能性もある。
そのため，分析対象から外すか否かは，注意して考える必要がある。

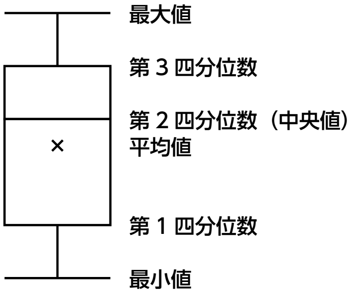
- **四分位数と箱ひげ図**

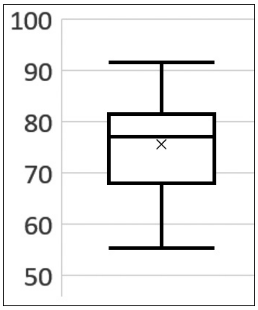
〔⑦　　　　　〕は，データを小さい順に並べたときに，データの数で 4 等分した時の区切り値のことである。下位から 25%，50%（真ん中），75%の位置にあるデータの値を，それぞれ第 1 四分位数（Q_1），第 2 四分位数（Q_2），第 3 四分位数（Q_3）という。

覚えておこう
第 2 四分位数は中央値でもある。データ数が偶数の場合は，中央の 2 つの値の平均値が第 2 四分位数となる。



〔⑧　　　　　　　　〕は，分布の概要を表した図で，最大値・最小値・四分位数の情報を表現したグラフのことである。箱ひげ図には，平均値の情報が含まれることもある。



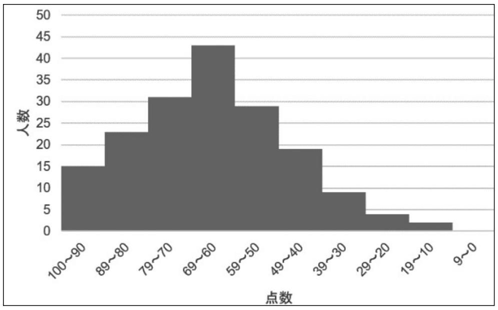


• 度数分布表とヒストグラム

〔⑨　　　　　　　　〕は，データをいくつかの階級に分け，階級ごとにそのデータ数（度数）を示して分布の様子をあらわした表である。

〔⑩　　　　　　　　〕は，度数分布表をもとに，縦軸に度数，横軸に階級をとって柱状のグラフであらわしたものである。

点数	人数
100～90	15
89～80	23
79～70	31
69～60	43
59～50	29
49～40	19
39～30	9
29～20	4
19～10	2
9～0	0
合計	175



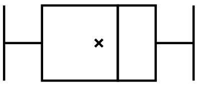
• クロス集計

〔⑪　　　　　　　　〕とは，2 つ以上のカテゴリ間でデータを比較するための集計方法のことである。各カテゴリ間の数や，その割合を計算し，それをまとめた表のことをクロス集計表という。

	A	B	C	D
1	満足度アンケート（クロス集計）			
2		満足	不満足	どちらともいえない
3	男性	55	9	11
4	女性	67	12	8
5	合計	122	21	19

覚えておこう

箱ひげ図は横向きで表現されることもある。



覚えておこう

表計算ソフトウェアでは，最大値は MAX 関数，最小値は MIN 関数で求めることができる。

覚えておこう

表計算ソフトウェアでは，合計値は SUM 関数で求めることができる。

重要

箱ひげ図では，第 1 四分位数から第 3 四分位数の距離（箱の長さ）の 1.5 倍以上箱から離れた値は外れ値とし，それを除いた数を最大値・最小値として図を描くこともある。

クロス集計は，アンケート結果の集計などに多く使われている。

例えば，次のクロス集計表は，内閣府が実施した，起業に対する意識調査アンケートの結果である。この表から，20歳代の起業意識は高く，年代を追うにつれて起業への関心が下がっていく傾向にあることが読み取れる。

(回答者数/回答対象者数)	全体	20歳代	30歳代	40歳代	50歳代	60歳代以上
周りのサポートがあれば起業してみたい	11.9%	13.9%	14.0%	12.8%	9.9%	7.7%
自分で起業してみたい	9.5%	12.5%	11.5%	10.0%	6.9%	4.9%
自分好みの会社をつくりたい	7.8%	10.1%	8.0%	7.8%	6.3%	5.8%
社会的起業をしてみたい	3.6%	4.2%	3.7%	3.4%	3.1%	3.4%
ユニコーン企業を立ち上げたい	1.8%	3.7%	2.6%	1.1%	0.7%	0.3%

【引用】内閣府 起業への関心アンケート

データの尺度と分析手法

集めたデータの〔⑫　　　　〕によっては，求めた値に意味がない場合もあるので，分析を行うときには注意が必要である。

復習
尺度とは，データの値を決める際の基準のこと。

	最頻値	最大値	中央値	最小値	平均値	分散
名義尺度	○	×	×	×	×	×
順序尺度	○	○	○	○	×	×
間隔尺度	○	○	○	○	○	○
比率尺度	○	○	○	○	○	○

例えば，徒競走の順位（1位，2位，3位…）は，順序尺度の1つである。

これは，順序関係に意味を持つが，値同士の間隔に意味はないため，そこから求められる平均値も意味を持たない。

まとめ

・代表的な統計量と表計算ソフトウェアの関数名

統計量	関数名
合計値	SUM 関数
平均値	AVERAGE 関数
中央値	MEDIAN 関数
最頻値	MODE 関数
最大値	MAX 関数
最小値	MIN 関数
分散	VARP 関数
標準偏差	STDEV.P 関数

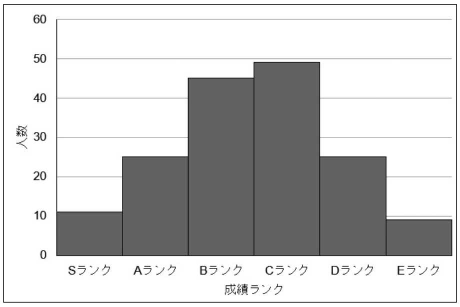
・**箱ひげ図**は，分布の概要を表した図で，最大値・最小値・四分位数の情報を表現している。

・**ヒストグラム**は，度数分布表をもとに縦軸に度数，横軸に階級をとって柱状のグラフであらわしたものである。

・**クロス集計**とは，アンケート結果などのデータを集計する場合に，回答データと設問を掛け合わせて集計する統計手法のことである。

Chapter2　実践問題

問１　以下のグラフは，成績ランクと人数をヒストグラムで表したものである。このヒストグラムを考察したア～エについて，正しければ○を，誤っていれば×を答えなさい。



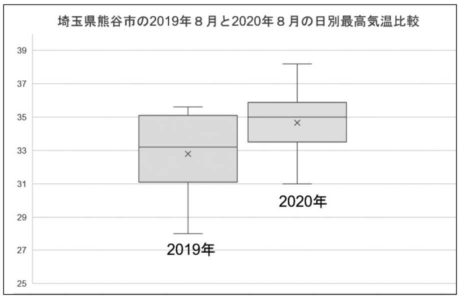
ア　A ランクの人数が最も少ない。

イ　S ランクから C ランクにかけて人数が多くなっていき，C ランクをピークに E ランクにかけては人数が少なくなっている。

ウ　全体の合計人数も容易に読み取ることができる。

エ　クロス集計表をもとに作られたグラフである。

問２　以下の箱ひげ図は，埼玉県熊谷市の 2019 年８月と 2020 年８月の日別最高気温を比較したものである。このグラフから読み取れることを 50 字程度で説明しなさい。



問３　以下のケンジ君とハナコさんの会話をもとに，クロス集計表を表計算ソフトウェアを使って完成させなさい。なお，合計値は関数を使って求めること。

ケンジ：今度のクラスマッチ大会で実施する競技について，サッカーがいいかドッチボールがいいか，男女別々にアンケートを取って集計したよ。
ハナコ：早速授業で習った内容を使って，クロス集計表にまとめたんだね！
ケンジ：男子のサッカー希望者は 82 人，ドッチボール希望者は 50 人になったよ。そして，女子のサッカー希望者は 65 人，ドッチボール希望者は 72 人になったよ。
ハナコ：希望者の数だけじゃなくて，サッカーとドッチボールそれぞれの希望者の合計，男女それぞれの人数の合計，総合計が表内にあるから傾向が分かりやすいね。
ケンジ：情報で習う内容って実際の生活でもすごく役に立つね！

問４　次の文章は，問３で作成したクロス集計表について考察したものである。（　）内の正しい文言をそれぞれ選びなさい。
男子と女子を合わせた合計人数は（269・137）人で，全体の合計としては（サッカー・ドッチボール）の希望者が多いが，（男子・女子）は，ドッチボール希望者の方が多い。

第9講　PART2　確認問題

問 1　以下の表は，表計算ソフトウェアで利用できる関数をまとめたものである。この表について，次の各問いに答えなさい。

関数	概要
（ア）	最頻値を求める
（イ）	標準偏差を求める
（ウ）	平均値を求める
（エ）	中央値を求める

(1)　空欄（ア）に入る語句として最も適切なものを，次の①～④のうちから一つ選びなさい。

①　AVERAGE　　②　MEDIAN
③　MODE　　④　STDEV.P

(2)　空欄（イ）に入る語句として最も適切なものを，次の①～④のうちから一つ選びなさい。

①　AVERAGE　　②　MEDIAN
③　MODE　　④　STDEV.P

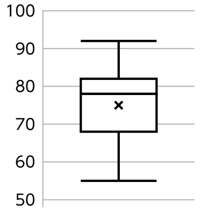
(3)　空欄（ウ）に入る語句として最も適切なものを，次の①～④のうちから一つ選びなさい。

①　AVERAGE　　②　MEDIAN
③　MODE　　④　STDEV.P

(4)　空欄（エ）に入る語句として最も適切なものを，次の①～④のうちから一つ選びなさい。

①　AVERAGE　　②　MEDIAN
③　MODE　　④　STDEV.P

問２　次のグラフの種類として最も適切なものを，次の①～④のうちから選びなさい。



①　散布図　　②　箱ひげ図
③　ヒストグラム　　④　レーダーチャート

問３　クロス集計に関する説明文として最も適切なものを，次の①～④のうちから選びなさい。

①　データの最大値・最小値・四分位数の情報を表現する方法で，分布の概要を読み取ることができる。
②　データをいくつかの階級に分け，階級ごとにそのデータの数を示して分布の様子を集計する方法
③　縦軸にデータ数，横軸に階級を柱状のグラフで表現する方法
④　２つ以上のカテゴリ間で，データを比較するための集計方法

第９講　データの活用

PART1　データの収集と整理

問１　③→①→②

問２　（ア）④　（イ）①　（ウ）③　（エ）②

問３　（ア）○　（イ）○　（ウ）×　（エ）○

問４　エ

PART2　データの分析

問１　（ア）×　（イ）○　（ウ）×　（エ）×

問２　2019 年に比べて 2020 年の方が最大値，最小値，平均値いずれも高く，暑い日が続いたと考えられる。

問３　省略

問４　269，サッカー，女子

PART3　データの解釈１

問１　(1) 省略　(2) 省略

(3) 散布図が右肩上がり，相関係数が 0.9 以上であることから強い正の相関があるといえる。

問２　（ア）×　（イ）×　（ウ）○

PART4　データの解釈２

問１　(1) 省略　(2) 31577 円

問２　（ア）×　（イ）×　（ウ）×

問３　④→②→①→③

第9講　PART2　確認問題　解答

問 1

(1) ③

解説：

最頻値を求めるのは MODE 関数である。

(2) ④

解説：

標準偏差を求めるのは STDEV.P 関数である。

(3) ①

解説：

平均値を求めるのは AVERAGE 関数である。

(4) ②

解説：

中央値を求めるのは MEDIAN 関数である。

問 2　②

解説：

分布の概要を表した図で，最大値・最小値・四分位数の情報を表現したグラフのことを，箱ひげ図と呼ぶ。

問 3　④

解説：

クロス集計は，2 つ以上のカテゴリ間で，データを比較するための集計方法である。

なお，①は箱ひげ図，②は度数分布表，③はヒストグラムの説明である。